CCCTC-Binding Factor Recruitment to the Early Region of the Human Papillomavirus Type 18 Genome Regulates Viral Oncogene Expression

Paris, Christian; Pentland, Ieisha; Groves, Ian; Roberts, David; Powis, Simon J; Coleman, Nicholas; Roberts, Sally; Parish, Joanna

DOI:
10.1128/JVI.00097-15

CCCTC-Binding Factor Recruitment to the Early Region of the Human Papillomavirus Type 18 Genome Regulates Viral Oncogene Expression

Christian Paris[2], Ieisha Pentland[1], Ian Groves[3], David C. Roberts[2], Simon J. Powis[2], Nicholas Coleman[3], Sally Roberts[1,#] and Joanna L. Parish[1,#]

[1]University of Birmingham, School of Cancer Sciences, Vincent Drive, Edgbaston, Birmingham B15 2TT, UK

[2]University of St Andrews, School of Medicine, North Haugh, St Andrews, Fife, KY16 9TF, UK

[3]University of Cambridge, Department of Pathology, Tennis Court Road, Cambridge, CB2 1QP, UK

[#]**Corresponding authors**

Dr. Joanna Parish, University of Birmingham, School of Cancer Sciences, Vincent Drive, Edgbaston, Birmingham, B15 2TT

Email: j.l.parish@bham.ac.uk

Tel: +44 (0)121 4158151

Dr. Sally Roberts, University of Birmingham, School of Cancer Sciences, Vincent Drive, Edgbaston, Birmingham, B15 2TT

Email: S.Roberts@bham.ac.uk

Tel: +44 (0)121 4147459

Abstract word count: 157

Summary of Importance word count: 142

Text word count: 5328

Abstract

Host cell differentiation-dependent regulation of human papillomavirus (HPV) gene expression is required for productive infection. The host cell CCCTC-binding factor (CTCF) functions in genome-wide chromatin organization and gene regulation. We have identified a conserved CTCF binding site in the E2 open reading frame of high-risk HPV types. Using organotypic raft cultures of primary human keratinocytes containing high-risk HPV18 genomes, we show that CTCF recruitment to this conserved site regulates viral gene expression in differentiating epithelia. Mutation of the CTCF binding site increases the expression of the viral oncoproteins E6 and E7, and promotes host cell proliferation. Loss of CTCF binding results in a reduction of a specific alternatively spliced transcript expressed from the early gene region concomitant with an increase in the abundance of unspliced early transcripts. We conclude that high-risk HPV types have evolved to recruit CTCF to the early gene region to control the balance and complexity of splicing events that regulate viral oncoprotein expression.

Importance

The establishment and maintenance of human papillomavirus (HPV) infection in undifferentiated basal cells of the squamous epithelia requires activation of a subset of viral genes, termed early genes. Differentiation of infected cells initiates expression of the late viral transcripts allowing completion of the virus life cycle. This tightly controlled balance of differentiation-dependent viral gene expression allows the virus to stimulate cellular proliferation to support viral genome replication with minimal activation of the host immune response, thus promoting virus productivity. Alternative splicing of viral mRNAs further increases the complexity of viral gene expression. In this study, we show that the essential host cell protein CCCTC-binding factor (CTCF), which functions in genome-wide chromatin organization and gene regulation, is recruited to the HPV genome and plays an essential role in the regulation of early viral gene expression and transcript processing. These data highlight a novel virus-host interaction important for HPV pathogenicity.

Introduction

Papillomaviruses are a highly diverse family of small DNA tumor viruses that specifically infect the mucosal and cutaneous epithelium. HPV types that infect the mucosal epithelium are sub-divided into low-risk and high-risk groups depending on their association with cancer development (1,2).

Following infection of cells in the basal layer of epithelium, the viral genome is amplified and maintained as a low copy episome (estimated to be between 10-200 copies per cell (3)). RNA polymerase-II-dependent transcription of the early proteins is initiated from the early promoter located upstream of the E6 ORF (P_{97} in HPV16 and P_{105} in HPV18 and 31) within the viral upstream regulatory region (URR). This drives expression of the E6 and E7 oncoproteins in the basal cells and stimulates continued cellular proliferation. The E7 gene products target the Retinoblastoma family of proteins pRb/p105 (4) and p107 (5), which control cell cycle entry in the basal layer. E7 also targets pRb2/p130 (6), which is highly expressed in the upper layers of the epithelium and prevents cell cycle re-entry (7). To circumvent increased p53 expression and cell cycle arrest arising from E7 expression, high-risk E6 protein binds p53 and targets it for degradation (8). By promoting cell cycle re-entry and delaying differentiation, E6 and E7 facilitate virus replication in cells that would normally have exited the cell cycle. In the upper epithelial layers, viral genome copy number rises, in part as a result of increased production of the viral E1 and E2 proteins (9,10). Increased E2 expression is thought to repress E6 and E7 production (11), stimulating cellular differentiation and subsequent activation of the differentiation-dependent late promoter (12). This allows production of transcripts encoding E1^E4, which promotes viral genome amplification (13), and the L1 and L2 capsid proteins (14). This intricate balance and control of early and late gene expression is essential for completion of the HPV life cycle.

All HPV transcripts are polycistronic. Alternative splicing and polyadenylation of transcripts

79 further regulate HPV early gene expression and increase the repertoire of expressed
80 proteins (14-17). Exactly how splicing of the early transcripts is regulated is not clearly
81 understood but suboptimal configuration of the 3' splice sites is thought to allow selection
82 between alternative splice acceptor sites (14). HPV16 also up-regulates splicing factors in
83 differentiating epithelium to support late transcript processing (18,19), highlighting the
84 ability of HPV to manipulate the host environment to control gene expression and co-
85 ordinate the differentiation-dependent life cycle.

86 CCCTC-binding factor (CTCF) is a ubiquitous host architectural protein that binds 10,000 to
87 50,000 sites within the human genome (20). Dynamic, three-dimensional organization of the
88 human genome by CTCF controls numerous genomic processes, including transcription,
89 genetic imprinting, chromatin insulation and gene splicing (21-25). These functions are
90 coordinated by CTCF through its ability to form long range interactions, bringing together
91 distant regulatory elements to control gene expression (26), or by forming a roadblock which
92 slows the transcription machinery and alters co-transcriptional RNA splicing (21). Due to the
93 highly complex and regulated nature of HPV gene expression and post-transcriptional
94 processing, we hypothesized that CTCF regulates differentiation-dependent HPV gene
95 expression.
96

Materials and Methods

Bioinformatics

The DNA sequences for each HPV type screened are defined in Table 2. Predictions for CTCF binding site were made with a combination of CTCF binding site databases (http://insulatordb.uthsc.edu/ and http://bsproteomics.essex.ac.uk:8080/bioinformatics/ctcfbind.htm), or using *Storm* analysis software. The position weight matrices (PWM) utilized by these analysis tools have been previously published (27-29).

Plasmids and antibodies

pUC19-HPV6b, pBR322-HPV11 and pBR322-HPV16 were a gift from E-M. de Villiers, DKFZ, Germany. pBR322-HPV31 was a gift from L. Laimins, Northwestern University, USA. pGEMII-HPV18 was a gift from F. Stubenrauch, University of Tübingen, Germany and was used as a template for site directed mutagenesis (QuikChange II XL; Agilent Technologies, USA) to create pGEMII-HPV18-ΔCTCF that contains three conservative nucleotide substitutions (C^{2993}→T, G^{3005}→A, T^{3020}→C) within the E2 coding region. The plasmid pDrive-SP6-His-CTCF was a gift from D. Farrar (University of Essex, UK) and encodes human CTCF protein with a 10x histidine tag at the N-terminus.

CTCF antibody was purchased from Active Motif (Belgium). FLAG M2 and anti-cytokeratin 1/10 8.60 antibodies were purchased from Sigma-Aldrich (UK). Anti-cytokeratin 5 D5/16 B4 was purchased from Boehringer Mannheim Biochemica (Switzerland) and loricrin AF62 from Covance Research Products (UK). BrdU and p130 were purchased from Becton Dickinson (UK). Cyclin B1 H-433, HPV18 E6 (G-7), p53 DO1 and GAPDH were purchased from Santa Cruz Biotechnology (USA). Phospho-histone H3 (S10) was purchased from Cell Signaling (USA) and HPV18 E7 (8E2) from Abcam (UK). All fluorescent secondary antibodies were purchased from Invitrogen (UK). Rabbit polyclonal anti-HPV16 E2 antibody was obtained from Prof.

Thierry (Singapore) (9), monoclonal anti E1^E4 1D11 (30) and rabbit polyclonal anti-E1^E4 r424 (31) were used to detect HPV18 E1^E4.

Electrophoretic mobility shift assay (EMSA)

DNA fragments were amplified with a forward primer containing an M13-overhang (Sequences available upon request) using Master Mix S (PeqLab, Germany). The products of the first PCR reaction were then amplified in a second PCR reaction using a FAM-labeled M13 forward primer. CTCF protein was produced in an *in vitro* transcription translation reaction using the TNT® SP6 High-Yield Wheat Germ Protein Expression System (Promega, UK).

2µl of FAM-labeled DNA was incubated with 1µl CTCF protein in a 10µl reaction containing 0.5% NP40, 50 mM KCl and 10mM Tris-HCl pH 7.5, 0.1µg/µl poly dIdC, 5% Ficol 400, 1mM PMSF and 0.1mM DTT. Samples were incubated at room temperature for 1hr before separation on a 4.5% native polyacrylamide gel. FAM fluorescence was imaged at 520nm using a Typhoon FLA7000 (GE Healthcare Life Sciences, UK).

Chromatin immunoprecipitation (ChIP)

ChIP assays were carried out using the ChIP-IT® Express Enzymatic ChIP Kit (Active Motif) following the manufacturer's instructions. Cells were fixed in 1% formaldehyde for 3 minutes at room temperature and nuclei released by 40 strokes in a tight dounce homogenizer. DNA was purified using a GenElute PCR Clean-up Kit (Sigma-Aldrich). ChIP efficiency was assessed by qPCR using SensiMix™ SyBr Mastermix (Bioline, London, UK) using an MXPro 3000 (Agilent Technologies). Primer sequences used are available upon request. C_T values were calculated at a constant threshold for each experiment and percent of input DNA calculated using the standard curve.

149

150 *Keratinocyte culture, transfection and organotypic raft culture*

151 W12 cells containing episomal HPV16 genomes were cultured as previously described (32).
152 The transfection of normal primary foreskin keratinocytes (HFK) from neonatal foreskin
153 epithelia (ethical approval number 06/Q1702/45) was performed in Dr. S. Roberts'
154 laboratory by Dr. J Parish as previously described (31,33). To eliminate donor-specific effects
155 2 donor lines were used; one produced as described above and one commercially available
156 HFK line (Clonetics™, Lonza Group Ltd, Basel, Switzerland). Emerging cell colonies were
157 pooled and expanded as previously described (34). Genomes were extracted from each line
158 and sequenced to ensure that the mutations were present in the mutant genome containing
159 lines. Organotypic rafts were prepared (31) and cultured for 14 days in E medium without
160 epidermal growth factor to allow cellular stratification. Sixteen hours prior to harvesting,
161 20μM BrdU was added to the growth medium. Rafts were then fixed in 3.7% formaldehyde
162 (Sigma-Aldrich) and paraffin embedded prior to sectioning (Propath Ltd, Hereford, UK).

163

164 *Cell growth assay*

165 1 x 10^5 terminally gamma-irradiated J2-3T3 fibroblasts were seeded to each well of 3x 12
166 well tissue culture microtitre plates and left to adhere. Wells were then seeded with $1x10^4$
167 HFK lines in triplicate. The growth of cells was measured at day 1, 3 and 5 following removal
168 of J2-3T3 fibroblasts by washing with EDTA and PBS. 500μl growth medium and 50μl CCK-8
169 reagent (Dojindo Molecular Technologies, Inc) were added to each well and the plate was
170 incubated at 37 °C for 2-4 hours. Absorbance was read at 450nm using an iMark™
171 microplate reader (Bio-Rad). Wells that contained J2 3T3 fibroblasts but not HFK were used
172 as a blank for each plate.

173

174 *Immunofluorescence*

Four µm sections of organotypic cultures were placed on poly-lysine-coated slides and incubated at 50°C for 30 min. One section from each raft was stained with haematoxylin and eosin for morphological analysis. Antigens were retrieved using an agitated low-temperature method, as previously described (35), following immersion in Histoclear™ (National Diagnostics, Yorkshire, UK). Slides were blocked with 20% heat-inactivated goat serum, 0.1% BSA in PBS for 1hr at RT. Primary antibodies were incubated on the slides overnight at 4°C. Incubation in secondary antibody was subsequently performed at 37°C for 1 hr. DNA was stained with Hoescht 33342 before mounting in Fluoroshield™ (Sigma-Aldrich). Microscopic analysis was performed in a Nikon E600 epifluorescent microscope and images captured using a Nikon DXM1200F digital camera.

Chromogenic in situ hybridization

Nuclei positive for HPV DNA amplification in raft sections were detected with a biotin-conjugated high-risk HPV DNA specific probe using Leica Bond-Max technology, as described by the manufacturer (Leica Microsystems, Milton Keynes, UK).

Episome copy number determination

Southern blotting was performed as described previously (36). For qPCR analysis, relative quantities of HPV18 genome in total DNA, amplified with primers 5'-TTATAGGCGAGCCCAAAAAC-3' and 5'-CCAATCTCCCCCTTCATCTAT-3', were normalized against the TLR2 locus at chromosome 4q32 using the Pfaffl comparative C_T method (37).

Transcript analysis

RNA was extracted from 14-day-old HFK raft cultures using RNA-STAT 60 (AMS Biotechnology Ltd, UK). 5µg of RNA was treated with 1 unit of RQ1 DNase (Promega) for

30min at 37°C, which was subsequently inactivated for 10min. Reverse transcription was performed using Tetro cDNA synthesis kit (Bioline). 2µl of cDNA was used for the amplification of HPV transcripts using the primers listed in Table 1. Products were separated by electrophoresis and the relative intensity of each product measured using Image J.

Statistical analysis

A two-tailed, unpaired student's T-test was used to determine statistical significance.

Results

Identification of CTCF binding sites in alpha-HPV genomes by bioinformatic analysis. CTCF binding sites in the genomes of low-risk HPV types 6b and 11 and high-risk HPV types 16, 18, 31 were predicted using open access databases and *Storm* analysis software (Table 2). These motif identification tools use a combination PWM previously described (27-29). As hypothesized, all of the HPV types tested were predicted to bind CTCF at multiple sites, although the number of predicted binding sites within different HPV types varied, ranging from six sites in HPV16 to eleven sites in HPV6b and 18. Numerous predicted binding sites clustered within the late gene region of all types studied. An additional site was identified in the E2 open reading frame (ORF) that was conserved in the high-risk but not in the low-risk viral types.

Verification of CTCF binding sites. To confirm our *in silico* analysis CTCF binding was assessed *in vitro* by EMSA. Approximately 200 bp DNA fragments containing the predicted binding motifs were incubated with CTCF protein (Figure 1A) and complexes separated by electrophoresis (Figure 1B). A region of the c-Myc promoter, previously shown to bind CTCF (38), and a fragment of the BPV1 genome not predicted to bind CTCF were included as controls. Fragments were also incubated with wheat germ extract alone and *in vitro* translated luciferase to control for non-specific binding of proteins. Fragments were tested a minimum of three times and the relative strength of binding in comparison to the c-Myc positive control DNA fragment estimated (Table 2 and Figure 2).

The binding maps presented in Figure 2 show conservation of CTCF binding between HPV types. All types contain a cluster of CTCF binding sites within the late gene region, ranging from 2 in HPV6b to 4 binding sites in HPV16. Furthermore, the conservation of one to two CTCF binding sites within (or close to) the E2 ORF of the high-risk HPV types was confirmed.

Binding in this region was not detected in HPV6b or 11 with fragments amplified from this region (Table 2). The conservation of CTCF binding sites between HPV types supports our hypothesis that CTCF recruitment is an important virus-host interaction in the HPV life cycle.

CTCF associates with HPV16 and HPV18 genomes. Next, we used HPV16 and HPV18 genome-containing cells to ascertain whether CTCF associates with the viral genome in cells. W12 cells, derived from a low-grade cervical squamous epithelial lesion, contain ~100 episomal HPV16 genome copies/cell (39,40) and HPV18 transfected human foreskin keratinocytes (HFKs) contain ~200 episomal HPV18 copies/cell (Figure 5B). CTCF association with the HPV genomes was determined by ChIP followed by qPCR. In both HPV16 and HPV18 genome-containing cells grown in monolayer, we noted a significant enrichment of CTCF binding within the E2 ORF, coinciding with the CTCF binding site conserved in high-risk HPV types but not in low-risk types (Figure 3). In contrast, we failed to detect CTCF binding to the late gene region in either HPV16 or HPV18 genome-containing model systems.

Loss of CTCF binding to the HPV18 genome does not alter episome establishment or proliferation of primary human foreskin keratinocytes. To assess the biological function of CTCF binding within the E2 ORF, mutations were introduced into the HPV18 genome to prevent CTCF binding (Figure 4A). Three nucleotide substitutions were introduced into the predicted binding site that did not alter the amino acid coding sequence of E2 (ΔCTCF HPV18). It should be noted that CTCF also has the potential to bind to the complementary DNA strand within this region (at the sequence 5' CACCACCTGGTGGT 3') although the mutations introduced would also affect binding at this site. We observed a near complete loss of CTCF binding to the ΔCTCF HPV18 sequence in EMSA confirming that the mutations prevented CTCF binding (Figure 4B). HFKs were transfected with recircularized wild type (WT) or ΔCTCF HPV18 genomes and immortalized lines established. To account for donor-

specific effects, cells from two independent donors were transfected and all downstream analyses were performed on both lines. No significant differences in cellular morphology (data not shown) or growth were observed between WT and ΔCTCF lines (Figure 5A). The physical state of the HPV genomes was determined by Southern blotting and qPCR. Both WT and ΔCTCF HPV18 lines were shown to contain episomal HPV genomes at a similar copy number of approximately 200 copies/cell (Figure 5B and C). Importantly, we demonstrated a 10-fold reduction in CTCF binding to ΔCTCF HPV18 genomes compared to WT (Figure 5D).

Loss of CTCF binding induces a hyperproliferative phenotype in organotypic culture. To assess the biological function of CTCF recruitment to the HPV18 genome in differentiating epithelium, WT and ΔCTCF HPV18 HFK lines were grown in organotypic raft culture. Formaldehyde fixed 'rafts' were paraffin embedded and sectioned. Sections were stained with hematoxylin and eosin to assess morphology (Figure 6A). As previously described, the WT HPV18 genome containing rafts were increased in thickness and mitotic cells were visible in the lower and upper suprabasal layers of the rafts in comparison to rafts derived from HFKs that did not contain HPV18 genomes (13). This phenotype was enhanced in ΔCTCF HPV18 rafts, which were consistently thicker indicating increased cellular proliferation. Alongside these experiments, viral genome amplification was assessed by chromogenic *in situ* hybridization (C-ISH). No consistent differences were observed in the number of cells with amplified HPV genomes between WT and ΔCTCF HPV18 rafts, demonstrating that CTCF recruitment has a minimal role in viral genome amplification (Figure 6A and B).

Increased S phase and G2 entry is caused by loss of CTCF binding. The increase in hyperproliferation in ΔCTCF HPV18 rafts could either be explained by delayed epithelial differentiation, or by increased S phase entry. To assess molecular differentiation, raft

sections were stained for markers of undifferentiated keratinocytes (keratin 5), early differentiation (keratin 1) and late differentiation (loricrin) alongside E1^E4, a marker of the productive phase of the HPV life cycle (Figure 6C). Expression patterns of keratin 5, keratin 1 and loricrin were similar between WT and ΔCTCF HPV18 organotypic cultures with keratin 5 confined to the basal and parabasal layers with some non-specific staining visible in the cornified layer of the epithelium; keratin 1 and loricrin expressed in the suprabasal and upper layers, respectively. However, keratin 1 and loricrin staining highlighted differences in the morphology of cells in the suprabasal and upper layers of the epithelium; rather than a flattening of these cells in the upper layers, as can be seen in the WT HPV18 sections, the cells appeared to maintain a rounded morphology. This difference in morphology is also visible in the hematoxylin and eosin stained sections shown in Figure 6A.

BrdU incorporation was used to assess cell cycle entry and cellular DNA replication. BrdU positive cells were confined to the basal layer in rafts derived from untransfected donor keratinocytes (Figure 7A). Increased S phase entry was observed in the basal and suprabasal layers of the WT HPV18 rafts, as previously reported (36). In contrast, ΔCTCF HPV18 rafts displayed a significant decrease in S phase entry in the basal layer as compared to WT. A decrease in the percentage of BrdU positive cells was observed in the lower suprabasal layers of ΔCTCF HPV18 rafts compared to WT, although this did not reach significance. In contrast a significant increase in the percentage of BrdU positive cells was observed in the upper suprabasal layers of the ΔCTCF HPV18 rafts (Figure 7B).

Raft sections were also stained for cyclin B1 and phospho-histone H3 (Ser10) (P-H3) as markers of G2 and mitotic entry, respectively (Figure 7C and E). In agreement with the BrdU incorporation analysis (Figure 7A and B), a decrease in cells positive for cytoplasmic cyclin B1 was observed in the basal layer of ΔCTCF HPV18 lines compared to WT. No significant difference was observed in the lower suprabasal compartment but an increase in cytoplasmic cyclin B1 in the upper suprabasal layers was noted (Figure 7C and D). In

313 contrast, there was no difference in the number of cells positive for P-H3 in WT and ΔCTCF
314 HPV18 structures (Figure 7E and F). Taken together, these data indicate that there is an
315 increase in cell cycle entry with a corresponding increase in S and G2 phases in the upper
316 layers of the epithelium of ΔCTCF HPV18 cells. The cells appear to arrest at G2 phase as an
317 increase in mitotic entry is not observed. These data provide evidence that loss of CTCF
318 binding within HPV18 E2 ORF leads to a delay in cell cycle exit and an enhanced
319 hyperproliferative phenotype.
320
321 **CTCF binding within the E2 ORF controls the expression of viral oncoproteins E6 and E7.**
322 The increased cell cycle entry and hyperproliferation observed in the organotypic raft
323 cultures derived from HFK lines maintaining ΔCTCF HPV18 genomes could be due to an
324 increase in the expression of E6 and E7 viral oncoproteins. Detection of these proteins by
325 immunostaining is not currently possible, therefore raft sections were stained with
326 surrogate markers; p53 as a marker for E6 expression and pRb family member p130 for E7
327 expression (8,41). Cells stained positive for p53 in WT HPV18 raft sections were apparent
328 throughout the epithelia as previously reported (42), albeit at a noticeably decreased level
329 than rafts derived from untransfected HFKs (Figure 8A). In contrast, p53 positive cells were
330 undetectable in rafts derived from ΔCTCF HPV18 lines (Figure 8A and B). This observation is
331 consistent with an increase in E6 protein levels in ΔCTCF HPV18 compared to WT, resulting
332 in a decrease in detectable p53 protein. Similarly, immunostaining with p130-specific
333 antibodies revealed significant differences between WT and ΔCTCF HPV18 rafts (Figure 8C
334 and D). In wild type HPV18 rafts, p130 positive cells were confined to the upper layers, as
335 previously shown (42), and in contrast to HPV negative HFK raft sections where cells stained
336 positive for p130 in the parabasal, lower and upper suprabasal layers. However,
337 immunostaining of p130 in the ΔCTCF HPV18 raft sections revealed an almost complete loss

of p130 positive cells in the upper layers, suggesting increased and prolonged expression of E7 protein in the ΔCTCF HPV18 rafts compared to WT.

Since p53 and p130 expression only provide an indication of E6 and E7 activity, we also quantified expression of early transcripts that have the potential to encode E6 and E7 by reverse transcriptase-PCR (RT-PCR). As expected, the relative abundance of unspliced E6E7 transcripts in ΔCTCF HPV18 raft cultures was significantly increased compared to WT (Figure 9A and B). E6E7 transcript levels were also measured by qPCR using the same primer set as described above and compared to the human RPLPO gene (Life Technologies). A ratio of E6E7 transcript to RPLPO transcript in HPV18 wild type and ΔCTCF rafts was calculated using the Livak $2^{\Delta\Delta_{CT}}$ method. Donor 1 was shown to have a 21.19 fold increase (±10.48 s.e.) and donor 2 had a 44.08 fold increase (±26.95 s.e.) in E6E7 transcript in the HPV18 ΔCTCF rafts compared to wild type. In addition, western blot analysis of protein extracts from raft cultures harvested at day 14 demonstrated a clear increase in E6 and E7 protein levels (Figure 9C), which was consistent in both donor lines.

The increase in E6E7 unspliced transcript could be due to an increase in the activity of the early promoter. HPV encoded E2 protein is known to repress the activity of this promoter and changes in E2 expression could affect early promoter activity (43-45). To determine whether the expression level of E2 protein was affected by the mutations introduced into the E2 ORF in the ΔCTCF HPV18 genome, E2 protein levels in raft lysates were detected by western blotting and no changes were observed (Figure 9C). Furthermore, immunostaining of sections showed E2 staining in the intermediate and upper layers of the WT raft cultures, with obvious cytoplasmic and nuclear localization. As previously described, E2 staining was not detected in the basal and lower suprabasal cells, presumably because E2 protein levels are below the level of detection (9). No staining was detected in the HPV negative HFK raft control demonstrating specificity of the antibody. An equal intensity of E2-specific signal was observed in the upper layers of wild type and ΔCTCF HPV18 rafts, although a delay in E2

expression was consistently observed in ΔCTCF HPV18 rafts compared to WT. This is presumably due to an expansion of the E2 negative mid-layers of the epithelium caused by increased E6 and E7 expression (Figure 9D). Together, these data confirm that steady state E2 levels in the raft cultures were not affected by the mutations introduced into the HPV18 genome. Collectively, these data demonstrate that CTCF recruitment to the conserved site within the E2 ORF is important in the regulation of viral oncoprotein expression in the differentiation-dependent life cycle through a mechanism that does not involve aberrant E2 protein expression.

CTCF controls RNA splicing of early viral transcripts. A diverse range of early transcripts is expressed from the HPV genome as a result of numerous alternative splicing events (14,15). Alterations in the splicing events that are important in early gene expression in HPV infections could have a dramatic effect on the expression of early proteins and their truncated forms (E6*I, E6*II and E6*III (14)). Given its previously described role in the control of RNA splicing (21), CTCF binding to the E2 ORF could affect splicing of the early transcripts and viral oncoprotein expression. To test this hypothesis, RNA was extracted from raft cultures harvested at day 14 and early transcripts amplified by RT-PCR with primer pairs that have been previously designed to identify the specific splicing events that occur within the early region of the HPV18 genome (14,15). Amplification with a 5' primer that anneals at nucleotide 121, upstream of the first splice donor site at nucleotide 233, and 3' primer that anneals at nucleotide 3517, downstream of the five splice acceptor sites in the early region of HPV18 at nucleotides 416, 2779, 3434, 3465 and 3506 (14,15) was used to detect any major splicing events that occur in the early region of the HPV18 genome. Amplification of RNA from WT HPV18 rafts resulted in two major products with some minor products visible (Figure 10A). As previously described (14), the two major products of 708 and 195 base pairs were identified by sequencing and shown to be spliced at 233^416 and

929^3434, and 233^3434, respectively (Figure 10B). Both of these products were consistently expressed in five raft cultures from each individual donor line of WT HPV18 HFKs. Of note, the 195 base pair product, spliced between nucleotide 233 and 3434, was significantly reduced and in some cases absent in the ΔCTCF HPV18 raft cultures (Figure 10A and C). This is in contrast to the increase in unspliced transcript in the ΔCTCF HPV18 rafts (Figure 9A and B). A significant reduction in production of the short mRNA species (233^3434 spliced product) could therefore result in the observed increase in unspliced E6E7 transcripts. Further analysis of viral transcripts revealed that splicing events at nucleotides 233^416 and 929^3434 were not altered by loss of CTCF binding (Table 3). These experiments demonstrate that loss of CTCF binding at position 2989 within the HPV18 genome results in a significant alteration in splice site usage, with specific loss of 233^3434 spliced products in the early transcripts expressed.

Discussion

This study aimed to identify CTCF binding sites within the genomes of various HPV types and to understand the function of CTCF in the virus life cycle. *In silico* predictions were used to identify potential CTCF binding sites, a high frequency of which bound CTCF *in vitro*. The relative position of many of the CTCF binding sites is conserved. A cluster of CTCF binding sites was identified in the late gene region of all of the HPV types tested and binding within the E2 ORF appears to be conserved in the high-risk types, indicating that recruitment of CTCF to this region is related to the ability of the virus to induce cellular transformation. This suggests that the recruitment of CTCF to these regions was an early evolutionary event and that CTCF is important for the virus life cycle. Furthermore, the frequency of CTCF binding sites within the genomes of the HPV types analysed in this study show an enrichment of sites compared to the frequency of binding sites within the human genome (20).

In contrast to the binding of CTCF within the E2 ORF in HPV16 and 18, CTCF recruitment within the late gene region was not detected in genome-containing cells. The conservation of the CTCF binding site cluster in the late gene region suggests that recruitment of CTCF to the late region is important for a defined point in the HPV life cycle. During submission of this manuscript, we became aware of a study by Metha *et al* in which CTCF was shown to associate with the sites within the L2 gene of the late gene region of HPV31. Loss of CTCF binding to the HPV31 L2 gene appears to prevent viral genome amplification (Metha K, Gunasekharan V, Satsuka A, Laimins L, submitted for publication). However, our data show that CTCF does not bind within the late gene region in HPV16 and HPV18 in cells grown in monolayer culture. It is possible that CTCF recruitment to this region is promoted by cellular differentiation and this is important for capsid protein expression or viral genome amplification. Differentiation-induced loss of CpG methylation in the late region of episomal HPV16 genomes has been reported (46). CpG methylation can negatively regulate CTCF

binding (47), making this method of regulation of CTCF recruitment to the late gene region of the HPV genome in differentiating epithelium a plausible hypothesis.

Several host cell proteins are recruited to the HPV genome to regulate transcriptional control. The binding of host cell transcription factors to sequences within the URR to control early gene transcription has been well characterized. Transcriptional regulators such AP1 (48), SP1 (49), TFIID (50), TBP (51), NF1 and Oct-1 (52) have defined binding sites within the URR of all HPV types analyzed. Many other transcriptional regulators are recruited by association with the E2 protein including Brd4 (53), TaxBP1 (54), p300 and CBP (55). In contrast, very few host or viral proteins have been shown to specifically bind to the HPV genome outside of the URR, although evidence of C/EBPβ, Oct-1 and YY1 binding to sequences upstream of the late promoter within the E7 ORF in HPV18 has been reported (56-58). The recruitment of CTCF to a binding site that exists within the E2 ORF is to our knowledge the first description of a cellular factor recruited to a specific binding site outside of the URR or late promoter regions to control viral gene expression.

Mutation of the CTCF binding site within the E2 ORF of HPV18 has highlighted an important function of CTCF in the HPV life cycle. Growth of cells in organotypic raft culture was affected by abrogation of CTCF binding and we noted a significant increase in cellular proliferation coupled with enhanced E6 and E7 protein expression. These data provide evidence that loss of CTCF binding within the E2 ORF enhances E6 and E7 expression in differentiating cells, prolonging the proliferative potential of cells in the mid- and upper-layers of the stratified epithelium. It is interesting to note that although we observed an increase in cell cycle entry in the ΔCTCF HPV18 raft cultures, we did not observe an increase in mitotic entry. One possible explanation for this is that the raft cultures were harvested at 14 days when the epithelia were fully differentiated. It is possible that an increase in mitosis occurs as the epithelium is developing and that in a fully differentiated epithelium, the cells are more likely to arrest in G2 than progress through mitosis. Importantly, there were no

discernable effects on the overall expression of E2, although expansion of the mid-layers of the epithelium resulted in an apparent delay in E2 expression. In addition, viral genome replication and amplification were unaffected, suggesting that E1 protein levels were unaffected. This suggests that HPV18 and perhaps other oncogenic HPV types have evolved to bind CTCF in this region to regulate balanced and controlled E6 and E7 expression in the context of a productive infection. Interestingly, CTCF does not appear to bind to the site within the E2 ORF in integrated sequences in HeLa cells (59), even though three copies of the binding site exist (60). It is possible that CpG methylation prevents CTCF binding to this site in HeLa cells, as previously reported (47,61) and it is tempting to speculate that the apparent loss of CTCF binding in integrated HPV18 genomes in HeLa cells contributes to the high E6 and E7 expression in these cells.

It should be noted that CTCF binding sites have been identified within the genomes of large DNA viruses such as Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV). Mutation of sites to prevent CTCF binding has demonstrated that CTCF determines latency in these viruses by blocking epigenetic silencing of latency-associated promoter elements and mediating long-range interactions within the viral genome (62-68). This, in part, is thought to be through CTCF-dependent regulation of nucleosome organization and control of RNA polymerase-II recruitment to the latency control region (69,70). Whether CTCF binding within the E2 ORF of the HPV genome directly controls E6E7 transcript production through similar mechanisms is currently being explored.

CTCF binding within the host genome controls co-transcriptional alternative splicing events by creating a roadblock to processing RNA polymerase-II and thereby promoting inclusion of weak upstream exons (21). We therefore analyzed splicing events that occur in the HPV early transcripts in differentiating epithelium and demonstrated a significant increase in the unspliced early transcript that encodes the E6 and E7 oncoproteins. In addition, the transcript spliced directly at 233^3434 was markedly reduced in ΔCTCF HPV18 rafts, while

the abundance of all other spliced products was unchanged. The transcript spliced at 233^3434 could be used as a template for translation of E6*II and E5 (14). Multiple transcripts that are abundantly and equally expressed in our WT and ΔCTCF rafts potentially encode E5 protein, making it unlikely that E5 expression is affected by loss of CTCF binding. Whether loss of E6*II expression contributes to the phenotype observed in our mutant HPV18 HFK rafts remains to be determined.

Our data suggest that CTCF recruitment to the E2 ORF binding site is a control mechanism for the expression of unspliced and alternatively spliced early transcripts in the HPV life cycle. It is interesting that the current model of CTCF-mediated splicing regulation predicts that DNA-bound CTCF pauses RNA polymerase-II progression and promotes the inclusion of weak upstream exons by allowing the splicing machinery more time to process the nascent RNA strand (21). Our data support a role for CTCF in directing splicing events, but suggests that the function of CTCF in this process is likely to be more complex than the current model predicts. In our physiologically relevant model system, loss of CTCF binding results in both increased levels of unspliced transcripts and a complex alteration of splice site usage upstream of the CTCF binding site. Further study of CTCF in the regulation of RNA processing will likely highlight novel functions of CTCF in gene expression regulation.

Acknowledgements

We thank Dr. Joseph Spitzer and his patients for the collection and donation of foreskin tissue. CP was supported by a PhD studentship funded by the University of St Andrews, School of Medicine. IP is supported by a Cancer Research UK (CRUK) PhD Studentship awarded to JLP and SR. IG and NC are supported by a CRUK Programme Award (13080) to NC. JLP is supported by a Royal Society University Research Fellowship (UF110010).

References

1. **Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride, AA.** 2012. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nuc Acid Res **41**:D571–D578.

2. **Clifford GM, Smith JS, Aguado T, Franceschi S.** 2003. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. Br J Cancer **89**(1):101–105.

3. **Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, Stanley MA.** 2012. The biology and life-cycle of human papillomaviruses. Vaccine **30**:F55–F70.

4. **Dyson N, Howley PM, Münger K, Harlow E.** 1989. The human papillomavirus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science **243**:934–937.

5. **Davies R, Hicks R, Crook T, Morris J, Vousden K.** 1993. Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. J Virol **67**(5):2521–2528.

6. **Zhang B, Chen W, Roman A.** 2006. The E7 proteins of low- and high-risk human papillomaviruses share the ability to target the pRB family member p130 for degradation. PNAS **103**:437–442.

7. **Roman A.** 2006. The human papillomavirus E7 protein shines a spotlight on the pRB family member, p130. Cell Cycle **5**:567–568.

8. **Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM.** 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell **63**(6):1129–1136.

9. **Xue Y, Bellanger S, Zhang W, Lim D, Low J, Lunny D, Thierry F.** 2010. HPV16 E2 is an immediate early marker of viral infection, preceding E7 expression in precursor structures of cervical carcinoma. Cancer Res **70**(13):5316–5325.

10. **Burnett S, Strom AC, Jareborg N, Alderborn A, Dillner J, Moreno-Lopez J, Pettersson U, Kiessling U.** 1990. Induction of bovine papillomavirus E2 gene expression and early region transcription by cell growth arrest: correlation with viral DNA amplification and evidence for differential promoter induction. J Virol **64**(11):5529–5541.

11. **Thierry F.** 2009. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. Virology **384**(2):375–379.

12. **Spink KM, Laimins LA.** 2005. Induction of the human papillomavirus type 31 late promoter requires differentiation but not DNA amplification. J Virol **79**(8):4918–4926.

13. **Knight GL, Pugh AG, Yates E, Bell I, Wilson R, Moody CA, Laimonis LA, Roberts S.** 2011. A cyclin-binding motif in human papillomavirus type 18 (HPV18) E1^E4 is necessary for association with CDK-cyclin complexes and G2/M cell cycle arrest of keratinocytes, but is not required for differentiation-dependent viral genome amplification or L1 capsid protein expression. Virology **412**(1):196–210.

14. **Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM.** 2011. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. J Virol **85**(16):8080–

8092.

15. **Kho E-Y, Wang H-K, Banerjee NS, Broker TR, Chow LT.** 2013. HPV-18 E6 mutants reveal p53 modulation of viral DNA amplification in organotypic cultures. PNAS **110**(19):7542–7549.

16. **Johansson C, Somberg M, Li X, Winquist EBOM, Fay J, Ryan F, Pim D, Banks L, Schwartz S.** 2012. HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. EMBO J **31**(14):3212–3227.

17. **Chow LT, Reilly SS, Broker TR, Taichman LB.** 1987. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. J Virol **61**(6):1913–1918.

18. **Mole S, Milligan SG, Graham SV.** 2009. Human papillomavirus type 16 E2 protein transcriptionally activates the promoter of a key cellular splicing factor, SF2/ASF. J Virol **83**(1):357–367.

19. **Mole S, McFarlane M, Chuen-Im T, Milligan SG, Millan D, Graham SV.** 2009. RNA splicing factors regulated by HPV16 during cervical tumour progression. J Pathol **219**(3):383–391.

20. **Chen H, Tian Y, Shu W, Bo X, Wang S.** 2012. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS ONE **7**(7):e41374.

21. **Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S.** 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature **479**:74–79.

22. **Phillips JE, Corces VG.** 2009. CTCF: Master Weaver of the Genome. Cell **137**(7):1194–1211.

23. **Bell A, West A, Felsenfeld G.** 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell **98**(3):387–396.

24. **Farrell CM, West AG, Felsenfeld G.** 2002. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol **22**(11):3820–3831.

25. **Awad TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, Neiman PE, Collins SJ, Renkawitz R, Lobanenkov VV, Filippova GN.** 1999. Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. J Biol Chem **274**(38):27092–27098.

26. **Marshall AD, Bailey CG, Rasko JEJ.** 2014. CTCF and BORIS in genome regulation and cancer. Curr Op Genet Dev **24**:8–15.

27. **Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES.** 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. PNAS **104**(17):7145–7150.

28. **Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B.** 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell **128**(6):1231–1245.

29. **Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, Brown GD, Marshall**

A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell **148**:335–348.

30. **Roberts S, Hillman ML, Knight GL, Gallimore PH.** 2003. The ND10 component promyelocytic leukemia protein relocates to human papillomavirus type 1 E4 intranuclear inclusion bodies in cultured keratinocytes and in warts. J Virol **77**(1):673–684.

31. **Wilson R, Ryan GB, Knight GL, Laimins LA, Roberts S.** 2007. The full-length E1^E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression. Virology **362**(2):453–460.

32. **Pett MR, Herdman MT, Palmer RD, Yeo GSH, Shivji MK, Stanley MA, Coleman N.** 2006. Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. PNAS **103**(10):3822–3827.

33. **Ruesch MN, F S, Laimins LA.** 1998. Activation of papillomavirus late gene transcription and genome amplification upon differentiation in semisolid medium is coincident with expression of involucrin and transglutaminase but not keratin-10. J Virol **72**:5106–5024.

34. **Wilson R, Fehrmann F, Laimins LA. 2005.** Role of the E1--E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. J Virol **79**(11):6732–6740.

35. **Watson RA, Rollason TP, Reynolds GM, Murray PG, Banks L, Roberts S.** 2002. Changes in expression of the human homologue of the Drosophila discs large tumour suppressor protein in high-grade premalignant cervical neoplasias. Carcinogenesis. **23**(11):1791–1796.

36. **Delury CP, Marsh EK, James CD, Boon SS, Banks L, Knight GL, Roberts S.** 2013. The role of protein kinase A regulation of the E6 PDZ-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18. J Virol **87**(17):9463–9472.

37. **Gray E, Pett MR, Ward D, Winder DM, Stanley MA, Roberts I, Scarpini CG, Coleman N.** 2010. In vitro progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. Cancer Res **70**(10):4081–4091.

38. **Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV.** 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol **16**(6):2802–2813.

39. **Alazawi W, Pett M, Arch B, Scott L, Freeman T, Stanley MA, Coleman N.** 2002. Changes in cervical keratinocyte gene expression associated with integration of human papillomavirus 16. Cancer Res **62**(23):6959–6965.

40. **Stanley MA, Browne HM, Appleby M, Minson AC.** 1989. Properties of a non-tumorigenic human cervical keratinocyte cell line. Int J Cancer **43**(4):672–676.

41. **Genovese NJ, Banerjee NS, Broker TR, Chow LT.** 2008. Casein kinase II motif-dependent phosphorylation of human papillomavirus E7 protein promotes p130 degradation and S-phase induction in differentiated human keratinocytes. J Virol **82**(10):4862–4873.

42. **Wang HK, Duffy AA, Broker TR, Chow LT.** 2009. Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. Genes Dev

23(2):181–194.

43. **Bernard BA, Bailly C, Lenoir MC, Darmon M, Thierry F, Yaniv M.** 1989. The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. J Virol **63**(10):4317–4324.

44. **Romanczuk H, Thierry F, Howley PM.** 1990. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. J Virol **64**(6):2849–2859.

45. **Tan SH, Gloss B, Bernard H-U.** 1992. During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. Nuc Acid Res **20**(2):251–256.

46. **Kalantari M, Lee D, Calleja-Macias IE, Lambert PF, Bernard H-U.** 2008. Effects of cellular differentiation, chromosomal integration and 5-aza-2'-deoxycytidine treatment on human papillomavirus-16 DNA methylation in cultured cell lines. Virology **374**(2):292–303.

47. **Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Myers RM, Stamatoyannopoulos JA.** 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res **22**(9):1680–1688.

48. **Thierry F, Spyrou G, Yaniv M, Howley P.** 1992. Two API sites binding JunB are essential for human papillomavirus type 18 transcription in keratinocytes. J Virol **66**:3740–3748.

49. **LI R, Knight JD, Jackson SP, Tjian R, Botchan MR.** 1991. Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell **65**:493–505.

50. **Tan SH, Leong LE-C, Walker PA, Bernard H-U.** 1994. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. J Virol **68**:6411–6420.

51. **Ham J, Steger G, Yaniv M.** 1994. Cooperativity in vivo between the E2 transactivator and the TATA box binding protein depends on core promoter structure. EMBO J **13**:147–157.

52. **O'Connor M, Bernard H-U.** 1999. Oct-1 activates the epithelial-specific enhance of human papillomavirus type 16 via a synergistic interaction with NF1 at a conserved composite regulatory element. Virology **207**:77–88.

53. **You J, Croyle JL, Nishimura A, Ozato K, Howley PM.** 2004. Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. Cell **117**(3):349–360.

54. **Wang X, Naidu SR, Sverdrup F, Androphy EJ.** 2009. Tax1BP1 interacts with papillomavirus E2 and regulates E2-dependent transcription and stability. J Virol **83**(5):2274–2284.

55. **Quinlan EJ, Culleton SP, Wu SY, Chiang CM, Androphy EJ.** 2013. Acetylation of conserved lysines in bovine papillomavirus E2 by p300. J Virol **87**(3):1497–1507.

56. **Wooldridge TR, Laimins LA.** 2008. Regulation of human papillomavirus type 31 gene expression during the differentiation-dependent life cycle through histone modifications and transcription factor binding. Virology **374**(2):371–380.

57. **Kukimoto I, Takeuchi T, Kanda T.** 2006. CCAAT/enhancer binding protein beta binds to and activates the P670 promoter of human papillomavirus type 16. Virology **346**(1):98–107.

58. **Gunasekharan V, Hache G, Laimins L.** 2012. Differentiation-dependent changes in levels of C/EBP repressors and activators regulate human papillomavirus type 31 late gene expression. J Virol **86**(9):5393–5398.

59. **Johannsen E, Lambert PF.** 2012. Epigenetics of human papillomaviruses. Virology. **445**:205–212.

60. **Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, Qiu R, Lee C, Shendure J.** 2013. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. Nature **500**(7461):207–211.

61. **Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV.** 2007. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem **282**(46):33336–33345.

62. **Tempera I, Klichinsky M, Lieberman PM.** 2011. EBV latency types adopt alternative chromatin conformations. PLoS Pathog **7**(7):e1002180.

63. **Chen HS, Martin KA, Lu F, Lupey LN, Mueller JM, Lieberman PM, Tempera I.** 2014. Epigenetic deregulation of the LMP1/LMP2 locus of Epstein-Barr virus by mutation of a single CTCF-cohesin binding site. J Virol **88**(3):1703–1713.

64. **Kang H, Wiedmer A, Yuan Y, Robertson E, Lieberman PM.** 2011. Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. PLoS Pathog **7**(8):e1002140.

65. **Chen H-S, Wikramasinghe P, Showe L, Lieberman PM.** 2012. Cohesins repress Kaposi's sarcoma-associated herpesvirus immediate early gene transcription during latency. J Virol **86**(17):9454–9464.

66. **Li D-J, Verma D, Mosbruger T, Swaminathan S.** 2014. CTCF and Rad21 act as host cell restriction factors for Kaposi's sarcoma-associated herpesvirus (KSHV) lytic replication by modulating viral gene transcription. PLoS Pathog **10**(1):e1003880.

67. **Tempera I, Wiedmer A, Dheekollu J, Lieberman PM.** 2010. CTCF prevents the epigenetic drift of EBV latency promoter Qp. PLoS Pathog **6**(8):e1001048.

68. **Chau CM, Zhang X-Y, McMahon SB, Lieberman PM.** 2006. Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. J Virol **80**(12):5723–5732.

69. **Kang H, Cho H, Sung GH, Lieberman PM.** 2013. CTCF regulates Kaposi's sarcoma-associated herpesvirus latency transcription by nucleosome displacement and RNA polymerase programming. J Virol **87**(3):1789–1799.

70. **Hilton IB, Simon JM, Lieb JD, Davis IJ, Damania B, Dittmer DP.** 2013. The open chromatin landscape of Kaposi's sarcoma-associated herpesvirus. J Virol **87**(21):11831–11842.

71. **Klenova EM, Nicolas RH, U S, Came AF, Lee RE, Lobanenkov VV, Goodwin GH.** 1997. Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in

697 SDS–PAGE and the size of the expressed protein is affected by the UTRs and sequences within
698 the coding region of the CTCF gene. Nuc Acid Research **25**:466–466473.

699

700 **Figure Legends**

701 **Figure 1: *In vitro* analysis of the association of CTCF with HPV genomes.** (A) Western blot
702 analysis of *in vitro* translated CTCF protein. Lysate from ID13 (mouse) cells known to express
703 CTCF was loaded as a positive control. A band running at approximately 140 kDa was
704 present in the ID13 cell lysate and a slightly smaller band was present in the *in vitro*
705 translated CTCF reaction. Human CTCF is an 82 kDa protein but runs at approximately
706 130kDa on SDS PAGE (71), whereas the mouse homologue is slightly larger. (B) An example
707 of an EMSA of CTCF binding to predicted BPV DNA fragments. DNA fragments were
708 amplified and labeled with FAM by PCR. Fragments were mixed with binding buffer only
709 (DNA), *in vitro* translated luciferase protein (-) or *in vitro* translated CTCF protein (+), and
710 protein-DNA complexes separated on a native acrylamide gel. Free DNA is indicated at the
711 bottom of the gel and Protein-DNA complexes near the top. Each fragment was tested a
712 minimum of three times and the combined results are shown in Table 2. Fragments from
713 the c-Myc locus (positive control), a region of the BPV-1 genome that is known not to bind
714 CTCF (negative control) and fragment 11 from HPV18 and fragments 1 and 10 from HPV31
715 are shown in the representative EMSA shown. 18_11 and 31_10 bound CTCF with medium
716 strength (50-75% binding compared to the c-Myc positive control) and 31_1 did not bind
717 CTCF *in vitro*.
718
719 **Figure 2: Summary of *in silico* predicted CTCF binding sites and in vitro analysis.** Graphical
720 representations of the HPV16, 18, 31, 11 and 6b genomes are shown. ORFs are indicated on
721 each genome (light grey). Predicted CTCF binding sites are represented by the black bars.
722 The hashed bars on the periphery of each genome highlight fragments tested by EMSA and
723 the dark grey bars on each genome indicate those fragments that bound CTCF *in vitro*.
724

Figure 3: Association of CTCF with HPV genomes. Chromatin extracted from (A) HPV16 positive W12 cells and (B) HPV18 positive HFKs was immunoprecipitated with control antibody (rabbit IgG for W12 and FLAG M2 antibody for HPV18 HFKs) or CTCF specific antibody. Co-precipitating DNA was analyzed by qPCR. The x-axes indicate the position in the HPV genome amplified and each data point represents the central point in each amplicon. A graphical representation of the HPV genome is shown above each data set, which has been linearized for ease of presentation. The CTCF binding sites verified by EMSA (Figure 1 and Table 2) are indicated (dark grey ovals). Binding efficiency was normalized to negative control antibody using the $\Delta\Delta C_T$ method. The data represent the mean and standard error of three independent repeats.

Figure 4: Mutation of the CTCF binding site at position 2989 in HPV18. (A) Wild type HPV18 sequence between nucleotides 2976 and 3035 showing the primary CTCF binding site starting at nucleotide 2989 and the secondary binding site in lowercase. The amino acid sequence of E2 protein encoded within this region is shown below the DNA sequence. The 3 conservative nucleotide substitutions created in the mutated ΔCTCF HPV18 genome (C➔T^{2993}, G➔A^{3002} and T➔C^{3020}) are indicated (*). (B) Abrogation of CTCF binding was assessed by EMSA. The CTCF binding region of the c-Myc locus (positive control) and a region of the BPV-1 genome that does not contain CTCF binding sites (negative control), the CTCF binding region in the E2 ORF in wild type and ΔCTCF mutant genomes were amplified and FAM-labeled by PCR. DNA fragments were mixed with binding buffer (DNA) alone or with *in vitro* translated luciferase (-) or CTCF (+) and complexes separated on a native acrylamide gel. In agreement with data presented in Table 2, CTCF bound strongly to the wild type HPV18 (18_3) fragment as compared to the positive control, however binding of CTCF to the ΔCTCF mutant fragment was severely disrupted.

751 **Figure 5: Creation of HPV18 wild type and ΔCTCF mutant human foreskin keratinocyte**
752 **lines.** HFKs established from two independent donors were transfected with WT or ΔCTCF
753 HPV18 genomes. (A) Analysis of growth kinetics using a CCK-8 metabolic assay. Cells were
754 seeded at equal density at day 0 and growth of each line was measured at days 1, 3, 5 and
755 7. The data show the mean and standard error of two independent experiments performed
756 in triplicate. (B) HPV18 genome copy number was determined by qPCR analysis of *Dpn*I
757 digested DNA extracted from each line using the Pfaffl comparative C_T method and
758 normalized against *TLR2* locus (37). Data show the mean and standard error of three
759 independent repeats (donor 1, p = 0.9; donor 2, p = 0.2). (C) HPV18 genome status was
760 determined by Southern blot from extracted DNA from donor 1 (1) and donor 2 (2)
761 transfected with either wild type (WT) or ΔCTCF mutant (ΔC) HPV18 genomes (OC, open
762 circle; L, linear; SC, supercoiled). DNA was linearized with *Eco*RI, producing a single band of
763 similar intensity running at approximately 8kbps, demonstrating maintenance of viral
764 episomes at a similar a copy number in each line. Digestion with *Bgl*II shows minimal
765 multimeric/integrated HPV genomes in all lines. (D) Abrogation of CTCF binding by mutation
766 of the CTCF binding site was determined by ChIP. Chromatin was either immunoprecipitated
767 with FLAG (negative control) or CTCF antibody and the percentage of bound HPV18 genome
768 was determined by qPCR with primers that flank the CTCF binding site at position 2989. A
769 significant decrease in CTCF binding was observed in ΔCTCF HPV18 as compared to wild
770 type (** p = 0.01). The data shown represent the mean and standard error of two
771 independent repeats performed in duplicate (Donor 1; Donor 2 showed a similar decrease
772 in CTCF binding).
773
774 **Figure 6: Morphology and differentiation of HPV18 ΔCTCF organotypic raft cultures.** (A)
775 Organotypic raft cultures of HFK, WT HPV18 and ΔCTCF HPV18 lines were fixed at day 14
776 and sections were stained with haematoxylin and eosin to assess morphology (upper
777 panels). Sections were stained by chromogenic in situ hybridisation (C-ISH) to qualitatively

assess viral genome amplification (lower panels). Brown nuclear staining is present in cells with amplified HPV18 genomes. Scale bar = 10 µm. (B) The number of cells positive for C-ISH in wild type and ΔCTCF HPV18 sections was counted in 10 fields of vision from sections of three independent raft cultures from each line (n = 30). The data are shown as the mean and standard error. (C) Sections were stained with antibodies specific for keratin 5 (green, upper panel), keratin 1 (green, middle panel) or loricrin (red, lower panel). Sections were counterstained with Hoechst to highlight the nuclei (blue) and E1^E4 antibody to highlight productive areas of each section (red in upper and middle panels (rabbit antibody r424), green in the lower panel (mouse antibody 1D11)). Scale bar = 10 µm.

Figure 7: Cell cycle entry in wild type and ΔCTCF mutant HPV18 genome containing organotypic raft sections. Sections were stained with (A) anti-BrdU (green), (C) cyclin B1 (red) and (E) phospho-histone H3 (Green). DNA stained with Hoescht to highlight the nuclei (blue). Representative sections of WT and ΔCTCF HPV18 genome containing HFK rafts are shown. The white arrows indicate the basal layer, and the lower suprabasal/upper suprabasal boundary is highlighted with a dashed line. Scale bar = 10 µm. The percentage of cells stained positive for (B) nuclear BrdU, (D) cytoplasmic cyclin B1 and (F) Phospho-histone H3 in the basal, lower suprabasal (parabasal and lower spinous) and upper suprabasal (upper spinous and granular) layers of 15 fields of view of 3 independent rafts (n = 45) from each donor was determined. The data represent the mean and standard error. (B) A significant reduction in BrdU incorporation is observed in the basal layer of ΔCTCF HPV18 lines (*** p = 0.002), a small reduction is observed in the lower suprabasal compartment that did not reach significance (p = 0.07) and a significant increase in BrdU incorporation is observed in the upper suprabasal layers of the ΔCTCF HPV18 rafts compared to wild type (*** p = 0.0002). (D) A significant reduction in cyclin B1 positive cells is observed in the basal layer of ΔCTCF HPV18 lines (* p = 0.04), no difference is observed in the suprabasal compartment and a significant increase in cyclin B1 positive cells is observed in the upper

layers of the ΔCTCF HPV18 rafts compared to wild type (*** p = 0.00006). (F) No significant differences in P-H3 positive cells were observed.

Figure 8: Analysis of p53 and p130 degradation in wild type and ΔCTCF HPV18 organotypic raft sections. (A) Sections were stained with p53-specific antibody (green) and DNA stained with Hoescht (blue). The white arrows indicate the basal layer, and the lower suprabasal/upper suprabasal boundary is highlighted with a dashed line. Scale bar = 5 μm. (B) The percentage of cells positive for nuclear p53 staining in the basal, lower suprabasal (parabasal and lower spinous) and upper suprabasal (upper spinous and granular) layers of 15 fields of view of 3 independent rafts (n = 45) from each donor was determined. The data represent the mean and standard error. A significant reduction in p53 positive cells is observed in all layers of rafts derived from the ΔCTCF HPV18 lines (*** $p < 0.0005$). (C) Sections were stained with p130-specific antibody (green) and DNA stained with Hoescht (blue). The white arrows indicate the basal layer, and the lower suprabasal/upper suprabasal boundary is highlighted with a dashed line. Scale bar = 5 μm. (D) The percentage of cells positive for nuclear p130 staining in the basal, lower suprabasal (parabasal and lower spinous) and upper suprabasal (upper spinous and granular) layers of 15 fields of view of 3 independent rafts (n = 45) from each donor was determined. The data represent the mean and standard error. A significant reduction in p130 positive cells is observed in all layers of rafts derived from the ΔCTCF HPV18 lines (*** $p < 0.001$).

Figure 9: Analysis of unspliced E6E7 transcript and protein expression in organotypic raft culture. RNA extracted from 14-day-old raft cultures was converted to cDNA and amplified between nucleotide 121 and 295. The products of this PCR reaction are unspliced early transcripts(14). Amplification of GADPH from the same samples is shown as a loading control. (A) Products were separated by electrophoresis and (B) quantified by densitometry using Image J. An increase in E6E7 transcript was observed in ΔCTCF HPV18 lines established

from individual donors (donor 1, * p = 0.03; donor 2, * p = 0.03). (C) Proteins extracted from raft cultures were analyzed by western blot. Fold increase in virus protein expression compared to wild type (normalized to GAPDH protein) is indicated below each membrane section. The images shown are representative of three technical repeats of lysates extracted from two independent donor lines. (D) E2 protein localization (red in merged image; DNA is blue) in raft sections of HFK, wild type and ΔCTCF HPV18 genome-containing lines. The images shown are representative of two independent raft cultures of each individual donor line. Scale bar = 10 µm.

Figure 10: Loss of CTCF binding causes aberrant splicing of early transcripts. RNA extracted from 14-day-old raft cultures was converted to cDNA and amplified between nucleotide 121 and 3517. (A) The products were gel purified and sequenced. A graphical representation of the identified products is shown in (B). (C) The 195bp product was quantified using Image J and relative amounts normalized to wild type levels for each donor. The data shown represent the mean and standard error of RNA extracted from 3 independent raft cultures from each donor (donor 1, *** p = 0.0008; donor 2, *** p = 0.0095).

Amplicon	Forward primer 5' to 3'	Reverse Primer 5' to 3'	Reference
121-295	ATCCAACACGGCGACCCTAC	GCAGCATGCGGTATACTGTCTCTA	(14), this study
121-3517	ATCCAACACGGCGACCCTAC	ACGGACACGGTGCTGGAA	(14)
E1F1/E4R	CAACAATGGCTGATCCAGAAG	AGGTCCACAATGCTGCTTCT	(15)

Table 1: Primers used for HPV18 transcript analysis

Class	Type	Predicted motif	Fragment Tested from	Fragment Tested to	Name	Motif sequence	Confirmed *in vitro*	Relative strength of binding
High risk	**HPV18**	843	754	943	18_1	ATTCCAGCAGCTGTTTCTGA	No	n.d.
	AY262282.1	1205	1102	1297	18_2	CCATTAGGGG	No	n.d.
		2989	2926	3117	18_3	AAACCACCAGGTGGTGCCAG	Yes	strong*
		3487	3381	3575	18_4	CGGTGAGGGG	No	n.d.
		3620	3527	3718	18_5	TTGCCTGTAGGTGTAGCTGC	Yes	medium
		4505, 4537	4440	4638	18_6 and 18_7	CGTCCCCCAGTGGT/GTAACAATAGATGGGTCTGT	Yes	two medium bands
		None	4947	5155	18_8	N/A	No	n.d.
		None	5045	5253	18_9	N/A	No	n.d.
		5473	5381	5577	18_10	CATACAGAGG	Yes	medium
		5767	5655	5850	18_11	CACCACCTGCAGGA	Yes	medium
	HPV16	1282	1216	1405	16_1	AACTCAGCAGATGTTACAGG	No	n.d.
	NC_001526.2	2915	2852	3049	16_2	TAACCACCAAGTGGTGCCAA	Yes	strong*
		5118	5000	5207	16_3	CGCCTAGAGG	Yes	weak
		6127	6051	6278	16_4	CCTATAGGGG	Yes	weak
		6514	6426	6600	16_5	GAACCACTAGGTGTAGGAA	Yes	weak
		6859	6772	6957	16_6	CTCCCCCAGGAGGC	Yes	weak
	HPV31	615	534	713	31_1	ATAACAGTGGAGGTCAGTT	No	n.d.
	J04353.1	885	804	1008	31_2	TGGGGAGGGG	No	n.d.
		1093	1029	1200	31_3	CATGCAGAGG	No	n.d.
		1277	1182	1374	31_4	AACGCAGCAGATGGTACAGG	No	n.d.
		2332	2230	2406	31_5	CAACCACTGGCTGATGCTAA	No	n.d.
		2412	2357	2531	31_6	AATGCACTAGATGGCAACC	Yes	strong*
		2853	2801	3015	31_7	TAACCACCAGGTGGTGCCAG	Yes	strong*
		None	2894	3093	31_8	N/A	No	n.d.
		5179	5077	5273	31_9	CCTTTAGGGG	Yes	strong
		6431	6354	6540	31_10	CTACACCTAGCGGC	Yes	medium
Low risk	**HPV6b**	1357	1251	1460	6b_1	CATACAGAGG	No	n.d.
	NC_001355.1	None	2801	3007	6b_2	N/A	No	n.d.
		None	2887	3101	6b_3	N/A	No	n.d.
		4789	4715	4913	6b_4	TGTGCAGGGG	No	n.d.
		5018, 4987	4913	5102	6b_5 and 6b_6	CTATCACTAGATGATACCA/CCTATAGAGG	No	n.d.
		5424	5317	5515	6b_7	GCAGCCACAAGAGGGTGCAT	Yes	strong
		6109	5995	6199	6b_8	CAGCCATTAGGTGTGGGTGT	No	n.d.
		6263	6179	6382	6b_9	CCCAAAGGGG	Yes	medium
		7205, 7256	7155	7380	6b_10 and 6b_11	CGAATAGAGG/CGTTTAGGGG	No	n.d.
	HPV11	1357	1295	1494	11_1	CATAGAGAGG	No	n.d.
	FR872717.1	None	2801	3003	11_2	N/A	No	n.d.
		None	2900	3104	11_3	N/A	No	n.d.
		4058	3930	4153	11_4	TGCAAAGGGG	Yes	medium
		4781	4709	4898	11_5	TGTGTAGGGG	No	n.d.
		4920	4844	5041	11_6	CCACCTGTGGAGGCCAGTG	Yes	weak
		5415	5330	5501	11_7	GCAGCCACTAGAGGGTGCAG	Yes	strong
		6310	6243	6428	11_8	GTTCCAACGGGGGGCAGTC	Yes	weak
		6635	6544	6738	11_9	GAGCCACTAGGTGTATGTA	Yes	weak, smear
		6979	6872	7074	11_10	CCTCCACCAAATGGTACACT	No	n.d.

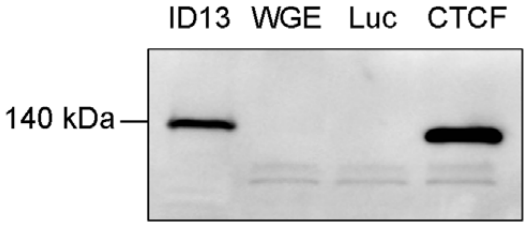
Table 2: Prediction of CTCF binding in various HPV types and relative *in vitro* binding affinity. The accession number of each HPV genome analyzed is indicated. The position in the viral genome of the first nucleotide of each predicted motif is given along with the specific fragment tested by EMSA. Where a fragment was tested that did not contain a predicted motif, no position is given (none). Each fragment tested is named by HPV type followed by the order of position in the genome starting at position 1 in the URR. The

sequence of each motif is given (N/A indicates that a motif was not predicted). Fragments were tested for binding *in vitro* by EMSA. The relative strength of binding to each fragment was assessed qualitatively by comparison to the proportion of c-Myc positive control DNA bound by CTCF in the same assay (weak, <50% binding; medium, 50-75% binding; strong, >75% binding; n.d. = none detected). All EMSA experiments were repeated at least three times and the strength of binding reflects the relative binding strength achieved in all repeats. *CTCF binding site within the E2 ORF that is conserved in all high-risk HPV types tested.

			Donor 1		Donor 2		
Primers	**Splice(s)**	**Product size**	**Fold change**	**p**	**Fold Change**	**p**	**Inferred ORFs**
121/295	Unspliced Early	175	1.65±0.27[a]	0.03	2.47±0.57	0.03	E6, E7
E1F1/E4R	929^3434	190	1.049±0.086	0.58	1.058±0.26	0.83	E1^E4, E5
121/3517	233^416, 929^3434	708	1.03±0.1	0.75	0.98±0.008	0.07	E6*I, E7, E1^E4, E5
121/3517	233^3434	195	0.37±0.087[b]	0.0008	0.59±0.11	0.009	E6*II, E5

Table 3: Analysis of splicing events in early transcripts produced in organotypic raft culture. RNA extracted from 14-day-old raft cultures was converted to cDNA and amplified with the indicated primer pairs. The fold change in transcript level compared to wild type HPV18 rafts is shown as the mean and standard error of three independent repeats. Significance (p) was calculated using a student's T-test. [a]Indicates a significant increase in expression levels. [b]Indicates a significant decrease in expression levels compared to wild type.

A



B

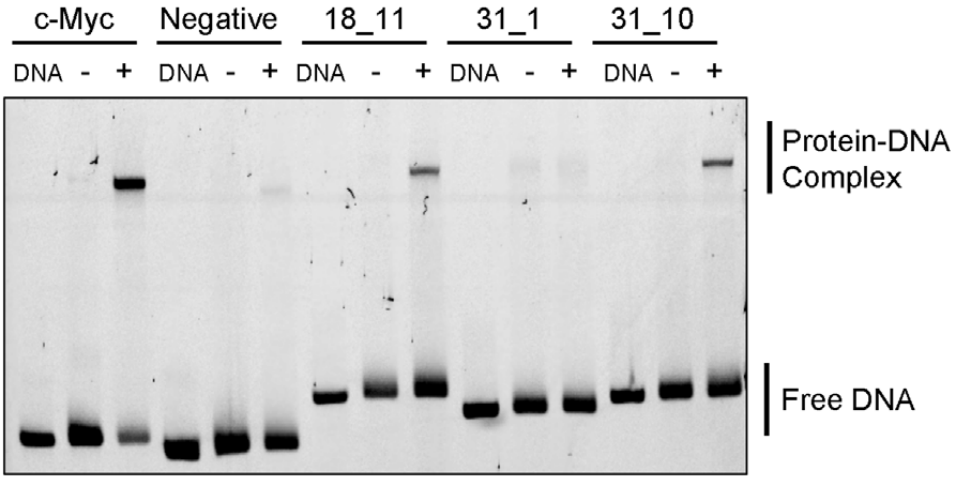


Figure 1

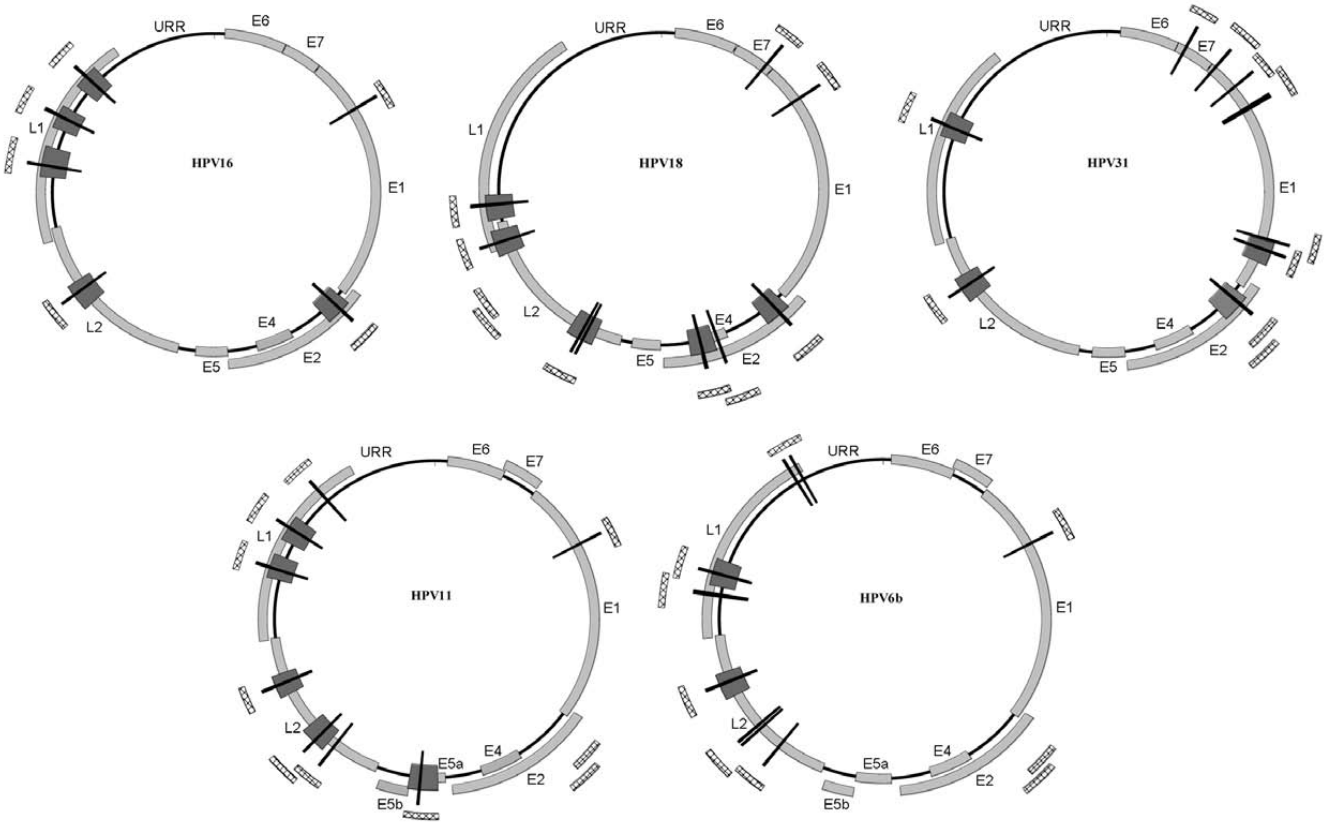


Figure 2

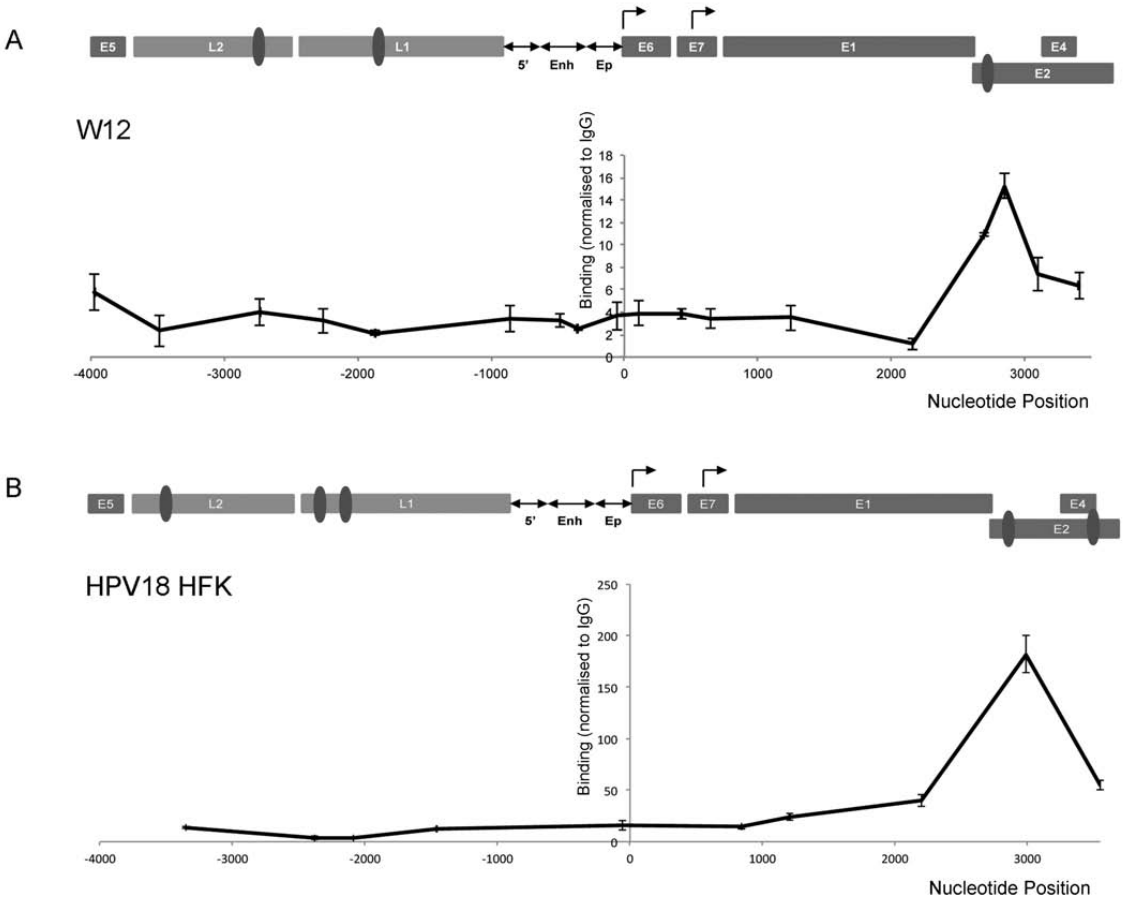
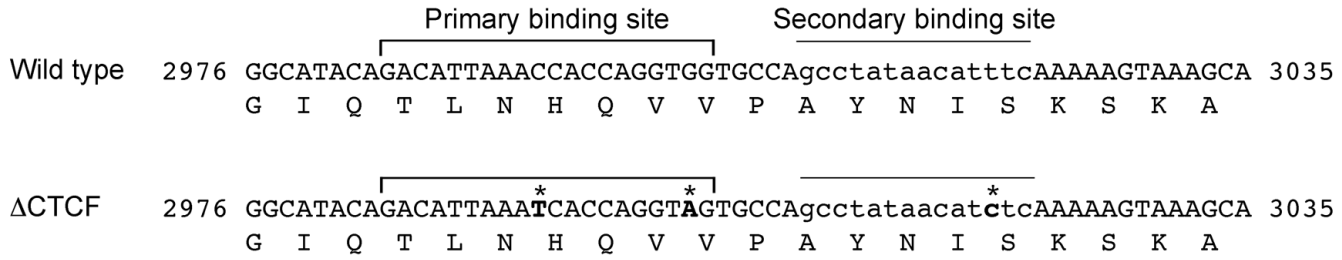


Figure 3

A



B

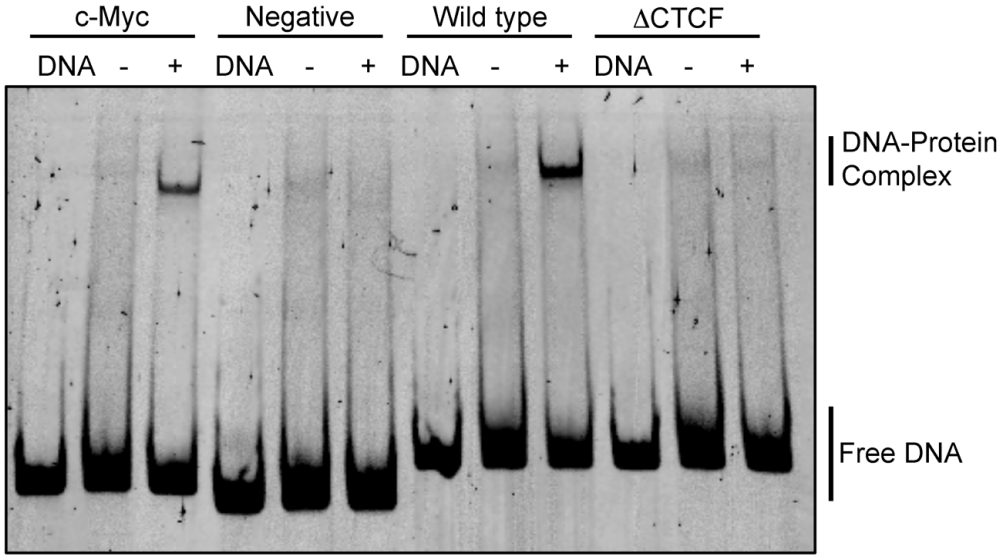


Figure 4

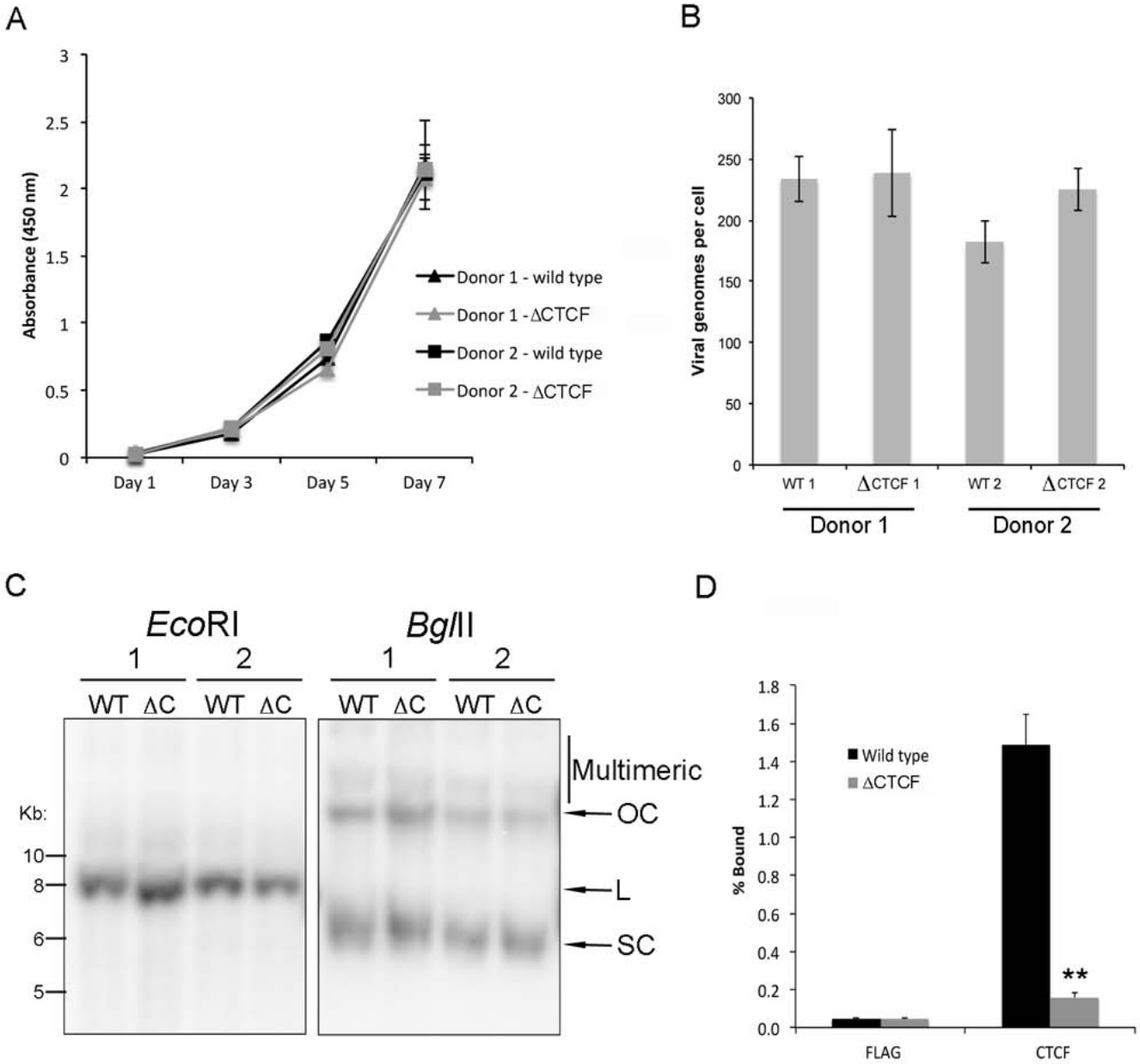


Figure 5

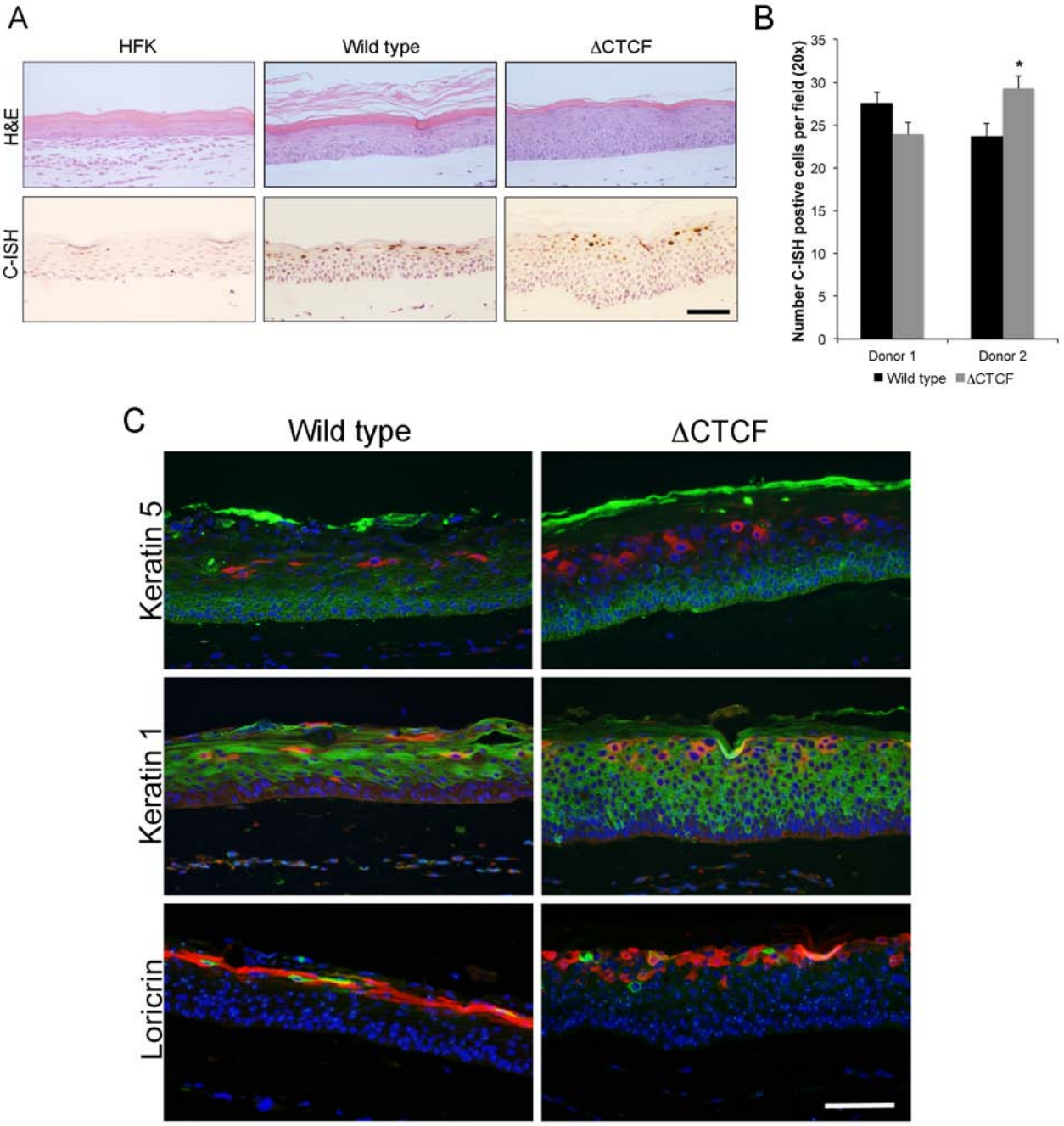


Figure 6

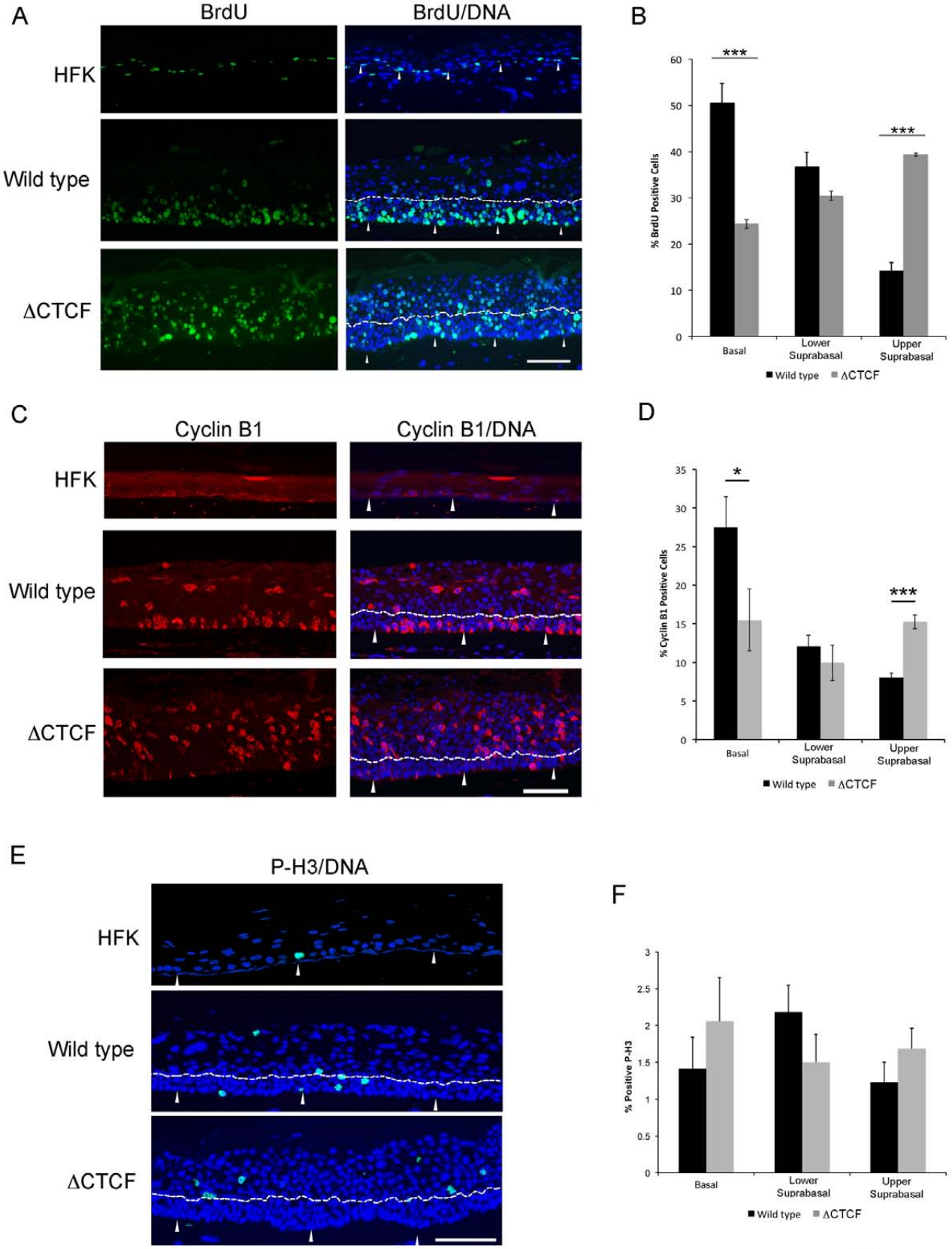


Figure 7

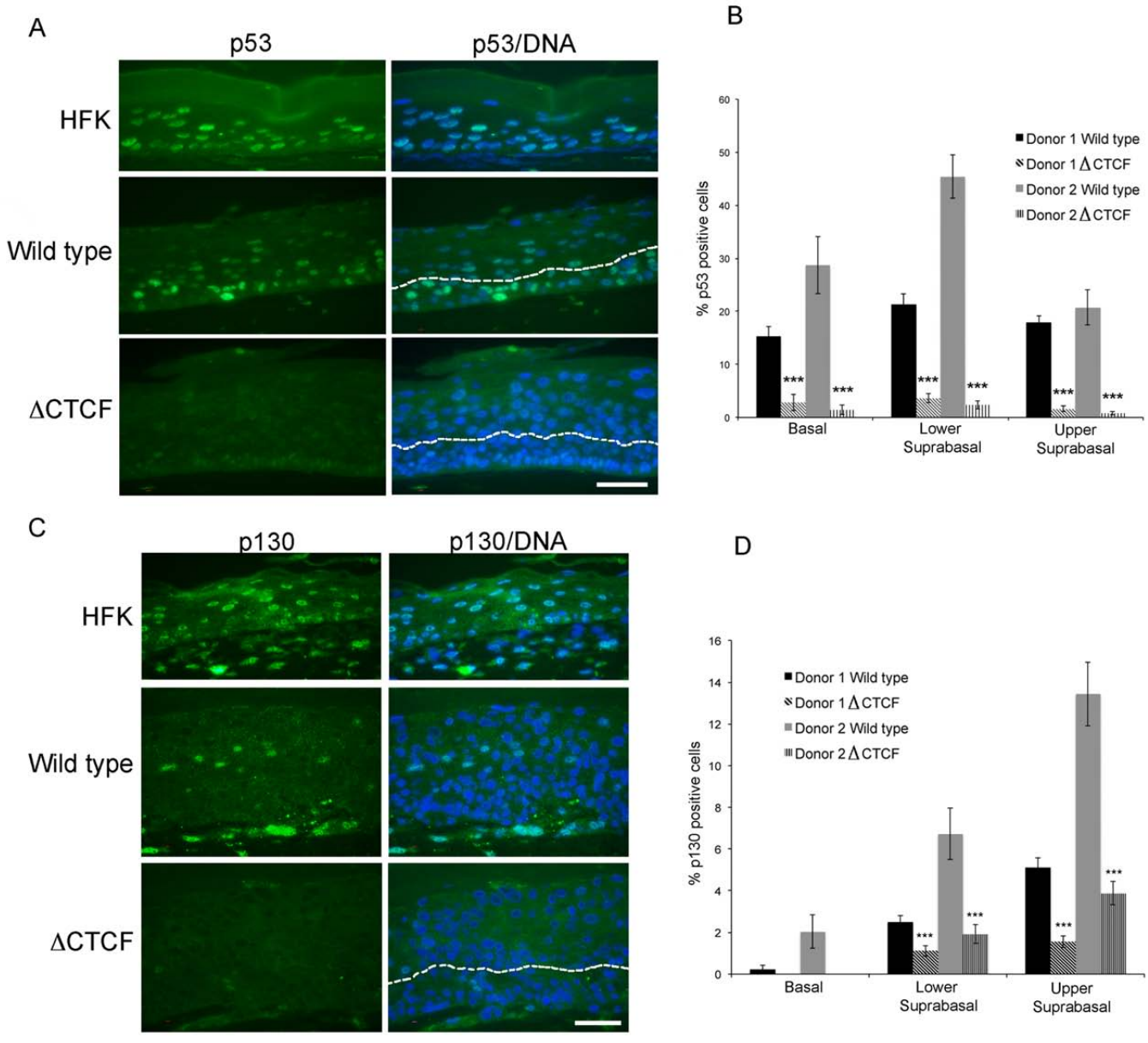


Figure 8

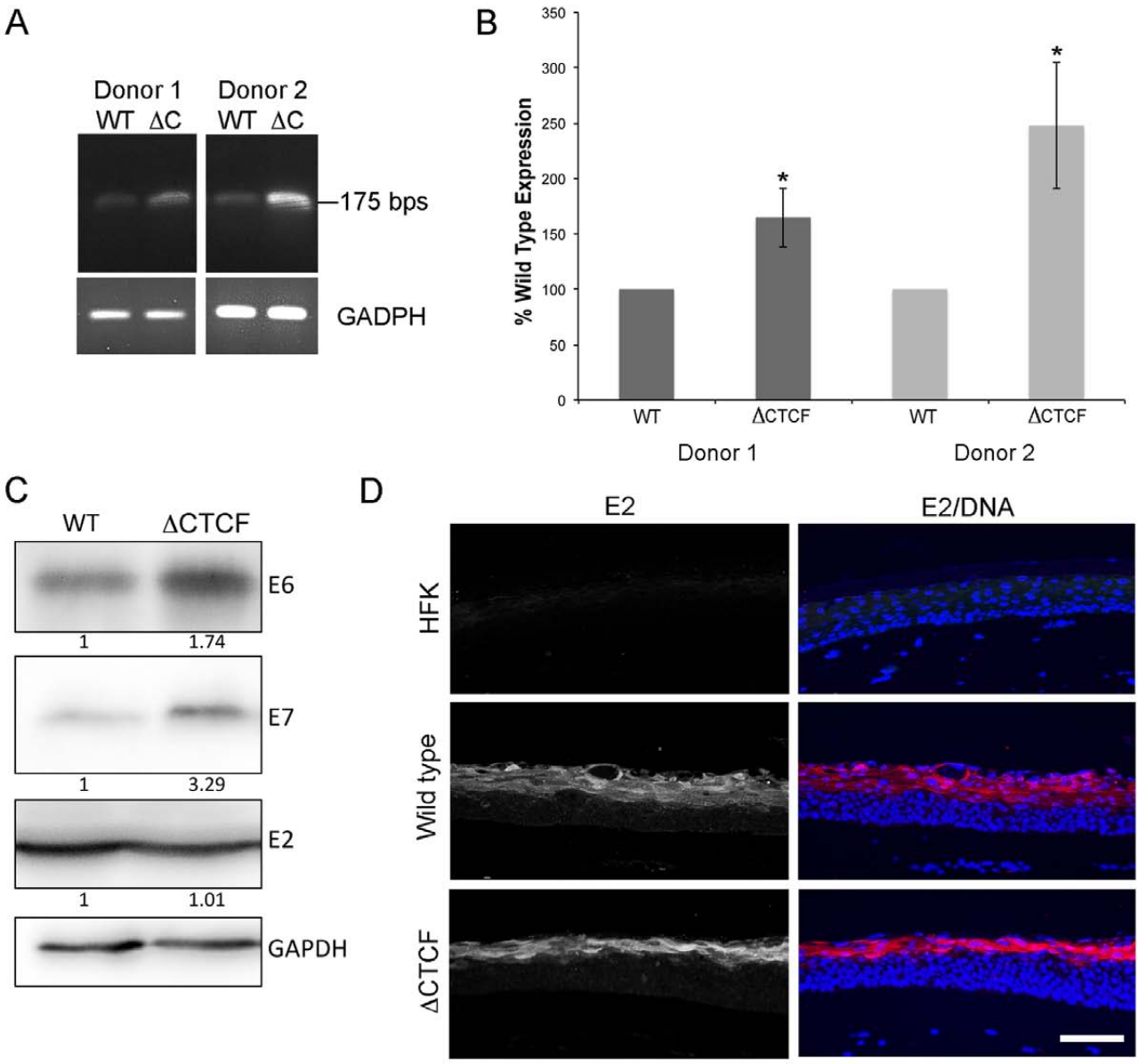


Figure 9

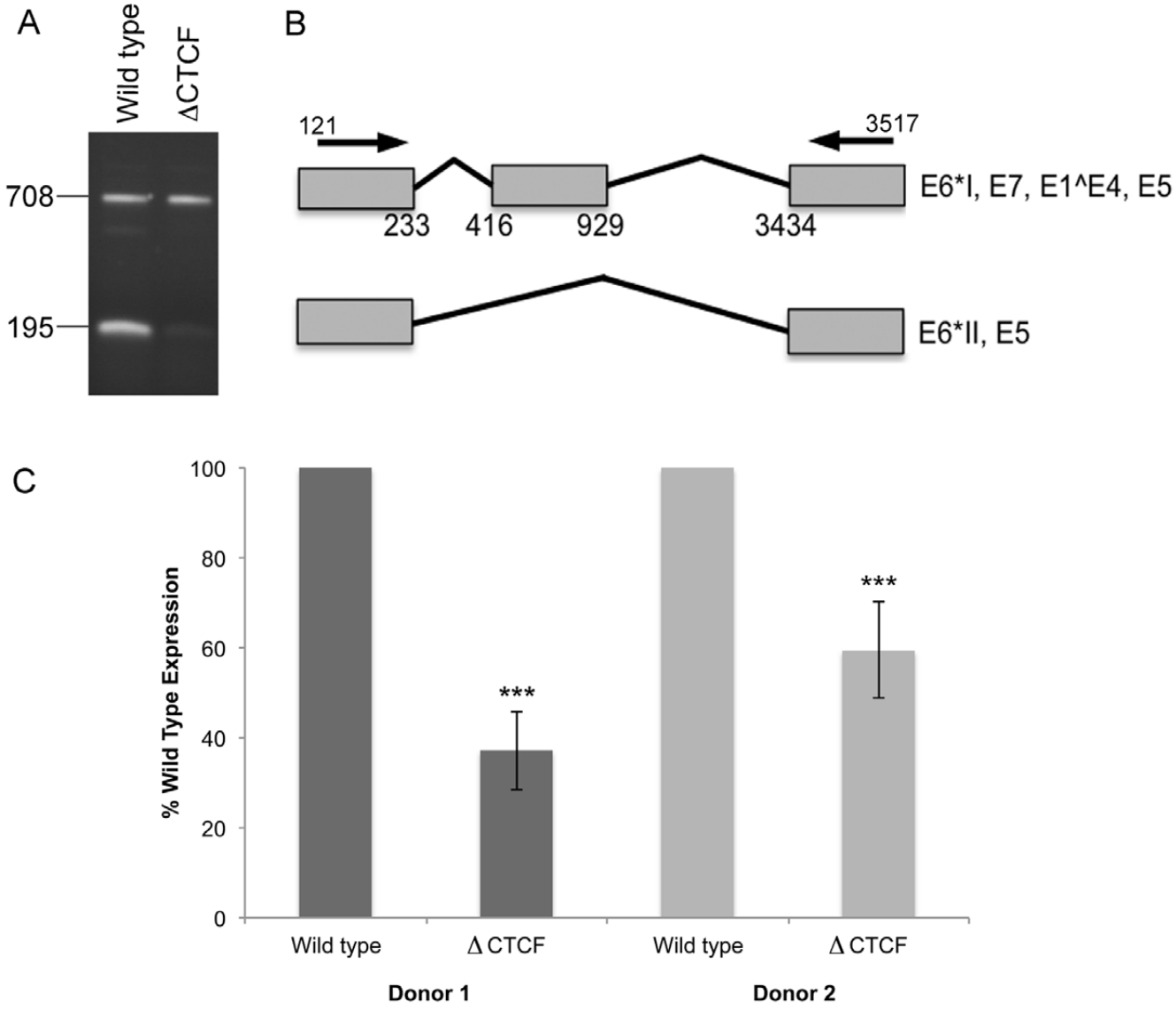


Figure 10